1. Generate summary statistics (including skewness and kurtosis) of the total number of patients discharged, average covered charges, average total payments, and average Medicare payments for Missouri and California states. Interpret and compare the results.

Discussion

	State	Metric	Mean	Median	SD	Skewness	Kurtosis
1	MO	total_discharges	37.62990	22.000	52.39025	5.662689	43.23767
2	MO	average_covered_charges	54434.82041	37385.587	57304.93369	5.269127	54.49727
3	MO	average_total_payments	12883.07195	9034.165	12850.89569	5.772021	59.71289
4	MO	average_medicare_payments	11183.39823	7777.956	11865.24729	5.621841	58.27115
5	CA	total_discharges	35.33716	20.000	57.95830	9.763183	166.95536
6	CA	average_covered_charges	109620.88407	72821.987	130445.98655	7.688773	117.20154
7	CA	average_total_payments	18508.20640	12663.073	20818.07824	7.113419	90.11586
8	CA	average_medicare_payments	16212.35581	11006.608	18822.38134	7.031908	88.91822

What is clear from the summary statistics is that while the total discharge numbers are similar the payment (covered, total, and medicare) per discharge is significantly larger for California compared to Missouri.

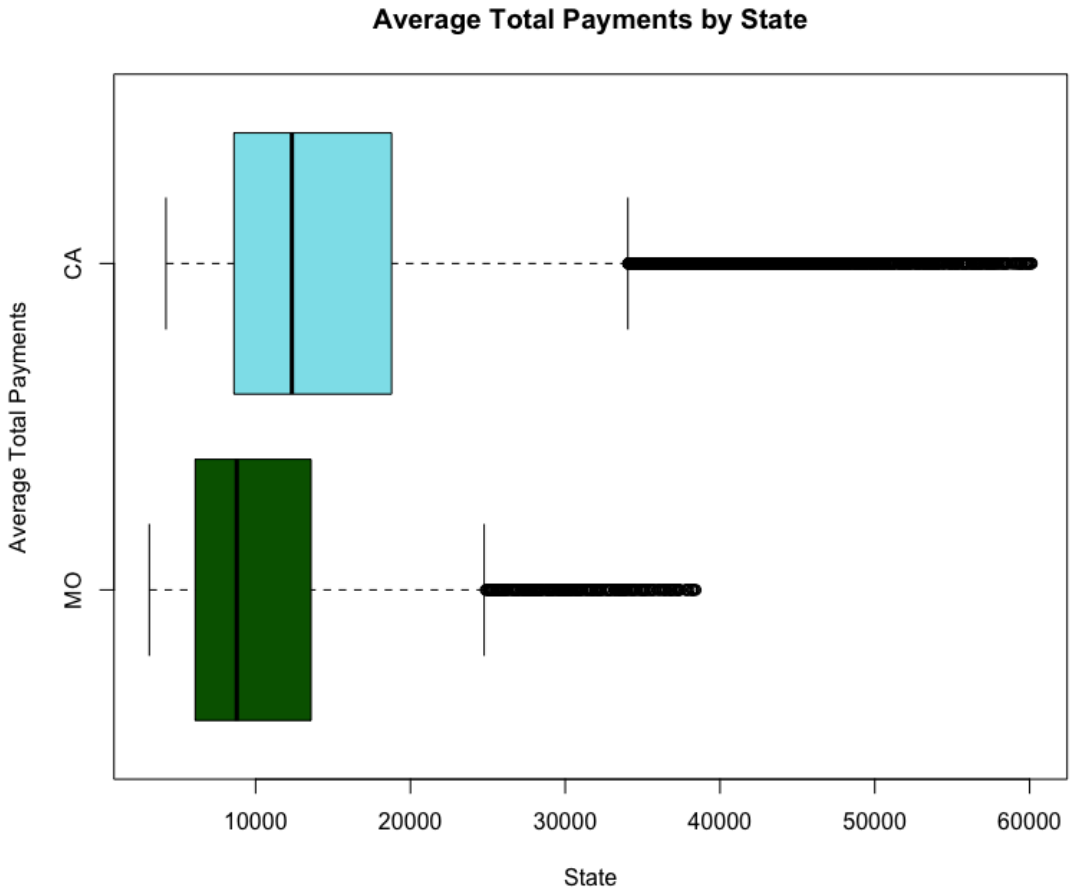
None of the statistics appear normal given that the Skewness is highly positive (indicating right-skewness) as well as high Kurtosis values indicating the presence of long tails, i.e. very flat distributions. Given these two characteristics it's likely the data does not consist of normal distributions.

The higher skew and kurtosis values for California support the higher payment averages and medians observed.

Thomas Jones
08206472

2. Generate box plots of the "average total payments" for Missouri and California, considering only those observations that fall under a 2-standard deviation of their corresponding means. Compare the plots in terms of median, spread, skewness, and outliers (if any).

Discussion



The box plot above displays the characteristics observed from question 1, namely the right-skewness of both sets of data as well as the higher median and average for California vs. Missouri. The long tail characteristics are clearly shown for both datasets and again, as observed, the high value of skewness and kurtosis for California as compared to Missouri is clearly indicated by how far the data extends past the IQR for California.

3. Generate the summary statistics, including minimum, maximum, and three-quartile values of "average covered charges" for the three urban states (California, New York, and Florida). Also, generate the summary statistics of the same variable for the three urban-rural mixed states (New Mexico, North Dakota, and Wyoming). Compare the results between urban states and urban-rural mixed states.

Discussion

	Type	Metric	Mean	Median	SD	Skewness	Kurtosis	Minimum	Maximum	First_Quartile	Second_Quartile	Third_Quartile
1	Urban	average_covered_charges	88081.83	59072.79	104935.20	7.711243	131.39533	3862.852	3427380.0	37570.77	59072.79	99839.48
2	Mixed	average_covered_charges	48460.74	33334.11	48360.28	4.070941	26.40469	6724.833	609885.6	22861.86	33334.11	53110.98

The data shows the same characteristics as the Missouri-California data. The inclusion of range and quartile data further details what would be shown in a histogram, i.e. a strong right skew with long tails.

4. Generate a data frame considering only those observations (rows) that are related to any kind of "CARDIAC" disease. Then, calculate the summary statistics, including average total discharges, median total discharges, average covered charges, and average total payments, for the following six states: California, Wyoming, Idaho, New York, Kansas, and Missouri. Comment on the summaries across six states.

Discussion

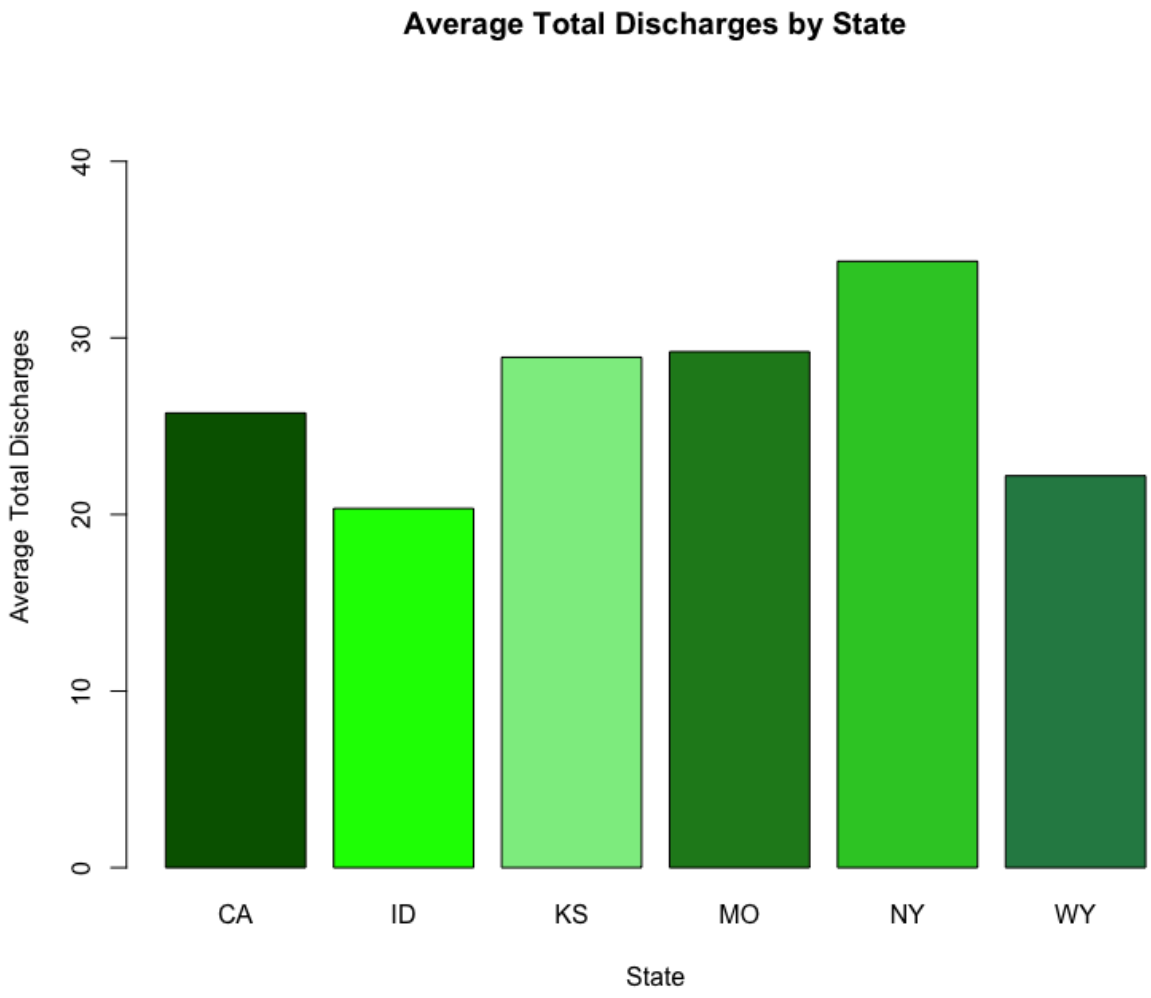
	state	avg_total_discharges	median_total_discharges	avg_covered_charges	avg_total_payments
1	CA	25.74487	19	168543.56	28170.09
2	ID	20.33898	18	87203.87	20516.17
3	KS	28.90323	22	85448.98	17797.34
4	MO	29.21575	22	88939.41	20971.78
5	NY	34.33898	23	110058.21	28241.73
6	WY	22.18750	18	55221.57	16981.16

While the data does not include it, just knowing that the populations of each of the states vary widely it is interesting that the average total discharges do not seem to be tightly correlated between the populations. That data might be worth further exploration but is out of scope of this assignment. The payments averages do appear correlative with urban dominated states, CA and NY, which could indicate that more densely populated areas have higher costs.

Thomas Jones
08206472

5. Generate a bar plot, placing six states, California, Wyoming, Idaho, New York, Kansas, and Missouri, on the x-axis and the average number of total discharges related to cardiac-related diseases on the y-axis. Comment on the plot.

Discussion



As discussed in (4) above, there does not seem to be a correlative relationship between population and number of discharges. While this would require additional exploration, it is interesting to note that MO and KS have similar rates as do ID and WY indicating that geographic location might be a possible area of correlation to explore.